# Categorical Perception of Speaker Identity

H.G. Tillman, L. Schiefer and B. Pompino-Marschall
*Munich, FRG*

## 1. Introduction

In natural speech communication an important role is played not only by linguistically defined categories, which determine one part of the phonetic structure of verbal utterances, but also by other aspects, such as the expression of emotion or the characteristics of the perceived individual speaker. We do not yet know very much about how the latter types of information are encoded in the acoustic speech signal.

In the framework of a research project on non-linguistic categories of perceived natural utterances we have focused our interest on the acoustic parameters of speaker identity. To determine the relevant parameters (or combination of parameters) it seems obvious to work with naturally produced material. The consequence is a situation which is much more complex than in the case of synthetic speech, where single parameters can be controlled easily.

To extract relevant parameters of speaker identity the experimental paradigm of categorical perception should be an interesting instrument. Our first experiments described below were undertaken in order to see whether there actually exists the possibility of categorical perception in the domain of speaker identification. It is our aim to apply this instrument to (more or less) complex test material the parameters of which are manipulated in a more sophisticated way.

## 2. Method

The starting points for the production of our test stimuli were the two German sentences 'Heute ist Donnerstag' and 'Aller Anfang ist schwer' uttered by two male (HGT/GK) and two female (LS/GH) speakers, respectively. These pairs of utterances were digitally recorded and processed in order to generate new stimuli placed at exactly equal acoustic distances between the two original ones. The linear interpolation of the acoustic parameters was achieved in the following way. First, the two original digital speech signals were segmented into voiced, fricative-voiceless and silent parts. Bursts were handled as short fricative segments. The duration of silent intervals was interpolated directly. Fricative segments, after appropriate time-warping, were interpolated in the amplitude-time-domain. In order to interpolate the voiced parts four parameters had to be manipulated: intensity contour, pitch contour, spectral (harmonic) structure and the duration, measured in numbers of pitch periods, each of a defined length. The duration of the respective voice parts of the initial utterances was determined by counting the number of pitch periods and summing their durations. The time-warping of the stimuli in the continuum between the original ones results from interpolating between these values.

Each individual pitch period (of both original utterances) was separately transformed into the frequency domain by computing the Discrete Digital Fourier Transform. Interpolation of the harmonic spectrum and retransformation with the computed values of $F_0$ and intensity by computing the Inverse Fourier Transform yielded the pitch periods of the new signals (a more detailed description of our stimulus generating programs is given in Simon (1983)). It should be added that as soon as the two original signals have been segmented properly by our speech editing system the experimenter is free to choose the number of stimuli to be computed between the two original ones. Even extrapolation is possible. The computed stimuli sound quite natural. Listening to the continuum itself, one perceives the change from one speaker to the other in discrete steps.

For our experiments the male continuum Cm consisted of 10 stimuli (including the original ones at the ends of the continuum), the female continuum Cf had 7 stimuli. The tapes for running the identification tests contained each stimulus 10 times in randomized order. There was an interstimulus interval of 4 s and a pause of 10 s after each 10 stimuli. For the discrimination tests pairs of two-step-neighbours of the respective continua were chosen as well as identical pairs. Thus a set of 26 pairs resulted for the male Cm and a set of 17 pairs for the female Cf material. The Cm-tape contained each pair 5 times in randomized order, and to produce the Cf tape each stimulus-pair was repeated 10 times. Within a pair of stimuli the pause was 500 ms, between the pairs themselves 4 s. Blocks of 10 pairs were separated by 10 s again.

In the identification tests the original utterances of the two speakers (HGT/GK, LS/GH were demonstrated 5 times, and the subjects were instructed that the utterances of these speakers had been computer-manipulated to varying degrees, and they were then asked to identify the speakers. In the discrimination tests they were asked to decide whether the utterances were identical or not.

## 3. Experiments

In our first experiment (Exp. I) we presented the stimuli of continuum Cm to 11 members of our institute who are very familiar with the voices of HGT and GK. The results of both identification and discrimination tests are presented in Fig. 1.

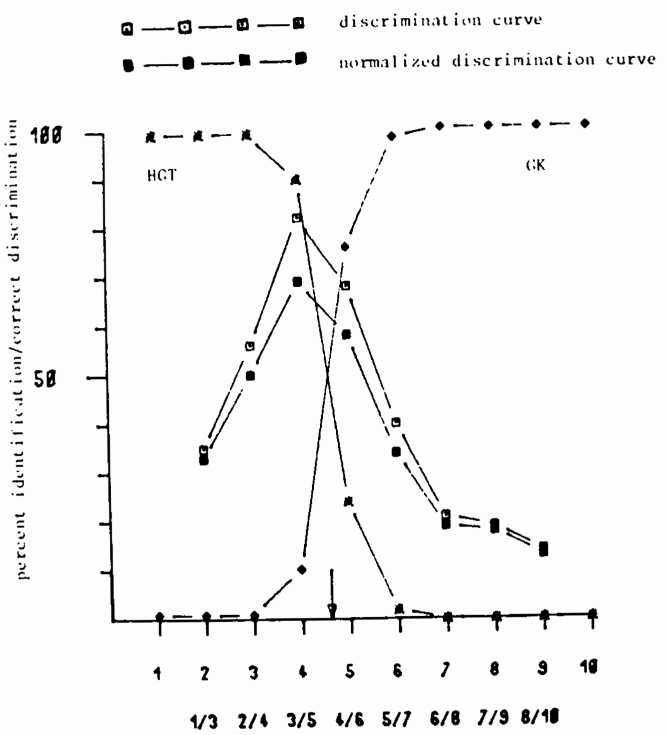

*Figure 1.* Results of Exp. I.

In Exp. II 15 subjects who had never heard the speakers of Cm before undertook first the discrimination test and then the identification test one week later (Cf. Fig.2)

In Exp. III the second continuum Cf was presented to the subjects of Exp. I. Again both speakers were known to them. (Cf. Fig. 3).

In Exp. IV the stimuli of Cf were presented to a group of 9 subjects to whom only one speaker, LS, was familiar. The discrimination test followed the identification test (Cf. Fig. 4)

In Exp. V the stimuli of Cf were presented to the subjects of Exp. II. This time the identification test was run first, and the discrimination test followed a week later. Again the speakers were not known to the subject. (Cf. Fig. 5)

## 4. Results and Discussion

The stimuli of continuum Cf and Cm were presented to three groups of listeners knowing either both, only one or none of the speakers, respectively. The results of the first group show clear categorical perception for both continua in the identification and discrimination tests (Exp. I and II, Fig. 1 and 3).

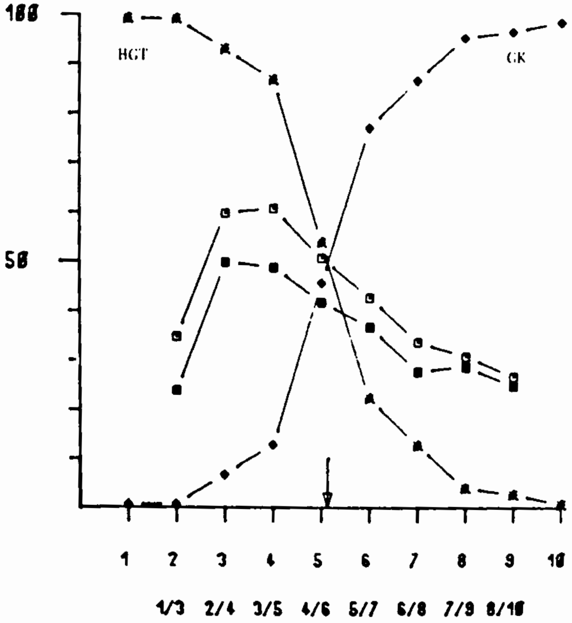

*Figure 2.* Results of Exp. II.

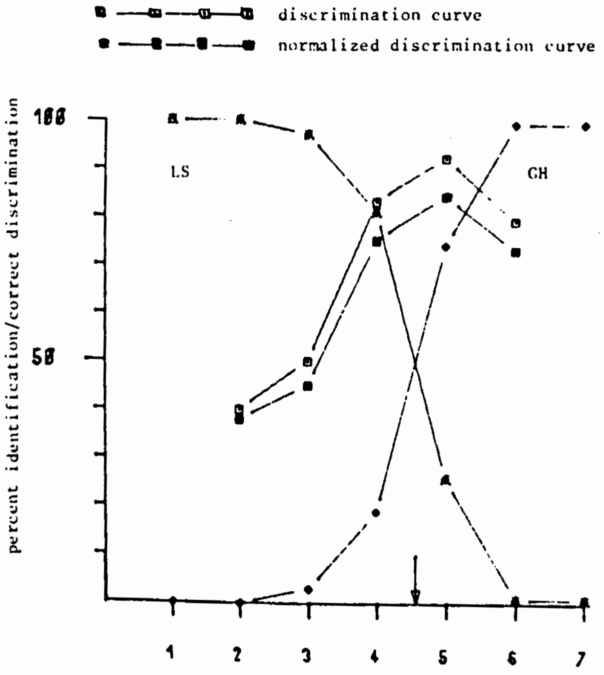

*Figure 3.* Results of Exp. III.

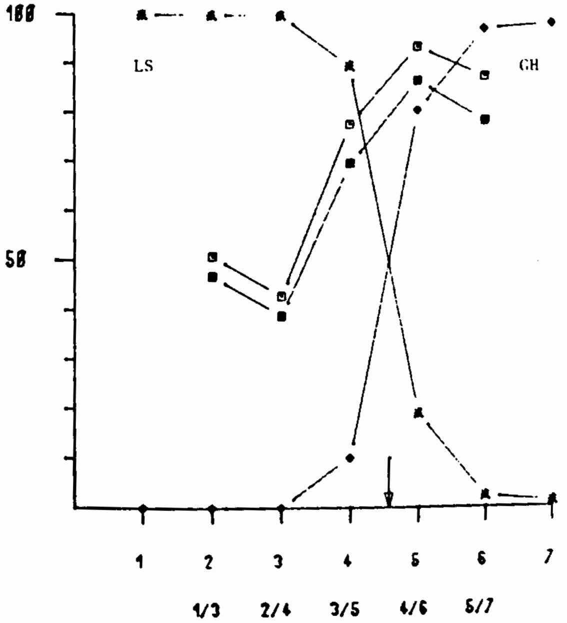

*Figure 4.* Results of Exp. IV.

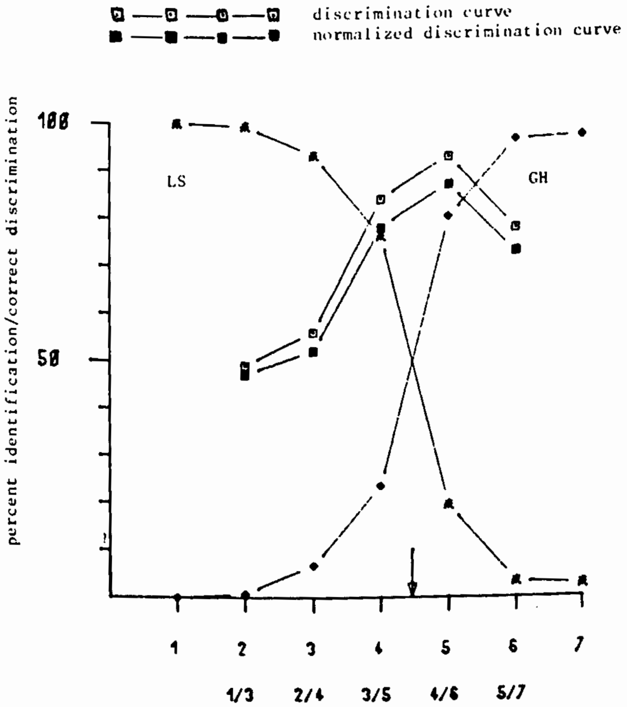

*Figure 5.* Results of Exp. V.

The second group obviously discriminated between the known and an unknown speaker and thus also produced categorical perception. (Exp. IV, Fig.4). No categorical perception was shown by the third group in Exp. II (Fig. 2) where the discrimination test was presented first. But this group reacted in a categorical manner when the discrimination test followed the identification test (Exp. V Fig. 5). Due to the ad-hoc complexity of the manipulated stimuli an interpretation of all details in the results cannot be given. Nevertheless some interesting facts should be mentioned. The discrimination curve (D-Curve) of group II for continuum Cf in Exp. IV (Fig.4) indicates that the difference between the original and the manipulated stimulus of the known speaker (i.e. stimulus 1/3, 3/1) leads to somewhat better discrimination. Another effect can be seen if one compares the original D-curves and those normalized according to the mean score for the respective identical pairs. Only in Exp. V (Fig. 5) do both D-curves have a nearly parallel form from the first until the last stimulus pair. In all other cases pairs of identical stimuli receive better 'same'- responses in the region of the identified speakers than in the region of the category boundary between the speakers. A third observation to mention is the dominance of speaker GK in Exp. I (Fig. 1), who wins 6:4 in the identification test, while on the other hand the D-curve shows better discrimination within the range of speaker HGT. This however correlates with the specific course of formants $F_4$ and $F_5$ in the utterance of speaker GK. In order to measure the influence of such different parameters of the phonetic form of utterances as $F_0$-contour, intensity contour, speech rate, different frequency regions of the spectrum etc., we are now preparing specific non-ad-hoc material which can be more easily manipulated in a systematic way. Finally it should be noted that also any 'artificial speaker' from the computed continua can be chosen as the starting point for the computation of a new continuum.

## References

Simon, Th. (1983). Manipulation of natural speech signals according to the speech parameters of different speakers. *Forschungsberichte des Instituts fuer Phonetik und Sprachliche Kommunikation der Universitaet Muenchen (FIPKM)* **17** (in press).

Tillmann, H.G. (1974). *Das individuelle Subjekt und seine persoenliche Identitaet im phonetischen Kommunikationsprozess.* Hamburg: Buske.

Tillmann, H.G., Simon, Th. (1983). Kategoriale und nichtkategoriale Komponenten in der Wahrnehmung bekannter und unbekannter Sprecher. *Forschungsberichte des Instituts fuer Phonetik und Sprachliche Kommunikation der Universitaet Muenchen (FIPKM)* **17** (in press).